

APR 25 2003

TECH CENTER 1800300



1600

ENTERED

RAW SEQUENCE LISTING

DATE: 04/22/2003

PATENT APPLICATION: US/10/037,677A

TIME: 14:28:34

Input Set : A:\GC560-D1-SEQLIST.txt

Output Set: N:\CRF4\04222003\J037677A.raw

```

4 <110> APPLICANT: Schellenberger, Volker
5     Liu, Amy D.
6     Selifonova, Olga V.
8 <120> TITLE OF INVENTION: Directed Evolution of Microorganisms
11 <130> FILE REFERENCE: GC560-D1
13 <140> CURRENT APPLICATION NUMBER: US 10/037,677A
14 <141> CURRENT FILING DATE: 2001-10-23
16 <150> PRIOR APPLICATION NUMBER: US 09/314,847
17 <151> PRIOR FILING DATE: 1999-05-19
19 <160> NUMBER OF SEQ ID NOS: 17
21 <170> SOFTWARE: FastSEQ for Windows Version 4.0
23 <210> SEQ ID NO: 1
24 <211> LENGTH: 741
25 <212> TYPE: DNA
26 <213> ORGANISM: Escherichia coli
28 <400> SEQUENCE: 1
29 atgaccgcta tgagcaactgc aattacacgc cagatcgttc tcgataccga aaccaccggt      60
30 atgaaccaga ttggtgcgca ctatgaaggc cacaagatca ttgagattgg tgccgttgaa      120
31 gtggtgaacc gtcgcctgac gggcaataac ttccatgttt atctcaaacc cgatcggctg      180
32 gtggatccgg aagcctttgg cgtacatggt attgccgatg aatttttgct cgataagccc      240
33 acgtttgccg aagtagccga tgagttcatg gactatattc gcggcgcgga gttggtgatc      300
34 cataacgcag cgttcgatat cggctttatg gactacgagt ttctggtgct taagcgcgat      360
35 attccgaaga ccaatacttt ctgtaaggtc accgatagcc ttgcggtggc gaggaaaatg      420
36 tttcccggta agcgcaacag cctcgatgcg ttatgtgctc gctacgaaat agataacagt      480
37 aaacgaacgc tgcacggggc attactcgat gcccgatcc ttgcggaagt ttatctggcg      540
38 atgaccggtg gtcaaacgct gatggctttt gcgatggaag gagagacaca acagcaacaa      600
39 ggtgaagcaa caattcagcg cattgtacgt caggcaagta agttacgcgt tgtttttgct      660
40 acagatgaag agattgcagc tcatgaagcc cgtctcgatc tgggtgcagaa gaaaggcgga      720
41 agttgcctct ggcgagcata a                                     741
43 <210> SEQ ID NO: 2
44 <211> LENGTH: 246
45 <212> TYPE: PRT
46 <213> ORGANISM: Escherichia coli
48 <400> SEQUENCE: 2
49 Met Thr Ala Met Ser Thr Ala Ile Thr Arg Gln Ile Val Leu Asp Thr
50 1           5           10           15
51 Glu Thr Thr Gly Met Asn Gln Ile Gly Ala His Tyr Glu Gly His Lys
52           20           25           30
53 Ile Ile Glu Ile Gly Ala Val Glu Val Val Asn Arg Arg Leu Thr Gly
54           35           40           45
55 Asn Asn Phe His Val Tyr Leu Lys Pro Asp Arg Leu Val Asp Pro Glu
56           50           55           60
57 Ala Phe Gly Val His Gly Ile Ala Asp Glu Phe Leu Leu Asp Lys Pro

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/037,677A

DATE: 04/22/2003

TIME: 14:28:34

Input Set : A:\GC560-D1-SEQLIST.txt

Output Set: N:\CRF4\04222003\J037677A.raw

```

58 65          70          75          80
59 Thr Phe Ala Glu Val Ala Asp Glu Phe Met Asp Tyr Ile Arg Gly Ala
60          85          90          95
61 Glu Leu Val Ile His Asn Ala Ala Phe Asp Ile Gly Phe Met Asp Tyr
62          100          105          110
63 Glu Phe Ser Leu Leu Lys Arg Asp Ile Pro Lys Thr Asn Thr Phe Cys
64          115          120          125
65 Lys Val Thr Asp Ser Leu Ala Val Ala Arg Lys Met Phe Pro Gly Lys
66          130          135          140
67 Arg Asn Ser Leu Asp Ala Leu Cys Ala Arg Tyr Glu Ile Asp Asn Ser
68 145          150          155          160
69 Lys Arg Thr Leu His Gly Ala Leu Leu Asp Ala Gln Ile Leu Ala Glu
70          165          170          175
71 Val Tyr Leu Ala Met Thr Gly Gly Gln Thr Ser Met Ala Phe Ala Met
72          180          185          190
73 Glu Gly Glu Thr Gln Gln Gln Gln Gly Glu Ala Thr Ile Gln Arg Ile
74          195          200          205
75 Val Arg Gln Ala Ser Lys Leu Arg Val Val Phe Ala Thr Asp Glu Glu
76          210          215          220
77 Ile Ala Ala His Glu Ala Arg Leu Asp Leu Val Gln Lys Lys Gly Gly
78 225          230          235          240
79 Ser Cys Leu Trp Arg Ala
80          245
82 <210> SEQ ID NO: 3
83 <211> LENGTH: 1164
84 <212> TYPE: DNA
85 <213> ORGANISM: Escherichia blattae
87 <400> SEQUENCE: 3
88 atgagctatc gtatgtttga ttatctggtt ccaaattgtga acttcttttg cccgggagcc 60
89 gtttctgttg ttggccagcg ctgccagctg ctgggggggta aaaaagccct gctggtgacc 120
90 gataaggggc tgcgcgccat taaagacggg gctgtcgatc agaccgtgaa gcacctgaaa 180
91 gccgcccgtg ttgaggtggg cattttcgac ggggtcgagc cgaacccgaa agacaccaac 240
92 gtgctcgacg gcctggccat gttccgtaaa gagcagtgcg acatgataat caccgtcggc 300
93 ggccggcagc cgcacgactg cggtaaaggc attggtattg cggccacca cccgggtgat 360
94 ctgtacagct atgccggtat cgaaacactc accaaccgcg tgcgcgccat tattgcggtc 420
95 aacaccaccg ccgggaccgc cagcgaagtc acccgccact gcgtgctgac taacaccaaa 480
96 accaaagtaa aattttgtgat tgtcagctgg cgcaacctgc cttccgtctc cattaacgat 540
97 ccgctgctga tgateggcaa gcccgccggg ctgaccgccg ccaccggtat ggatgccctg 600
98 acccacgcgg tagaggccta tatctccaaa gacgccaacc cggttaccga tgcctctgct 660
99 attcaggcca tcaaaactgat tgccaccaac ttgcgccagg ccgtcgccct ggggaccaac 720
100 ctcaaagccc gtgaaaacat ggctgcgcc tctctgctgg ccgggatggc ctttaacaac 780
101 gccaacctgg gctatgttca cgccatggct caccagctgg gcggcctgta cgacatggcc 840
102 cacggggtgg cgaacgcggg cctgctgccc catgtctgcc gctataacct gattgccaac 900
103 ccgaaaaaat ttgccgatat cgccaccttt atgggggaaa acaccaccgg tctttccacc 960
104 atggacgcag cggagctggc catcagcgcc attgcccgtc tgtctaaaga tgtcgggatac 1020
105 ccgcagcacc tgcgtgaact gggggtaaaa gaggccgact tcccgtacat ggcagaaatg 1080
106 gccctgaaag acggcaacgc cttctctaac ccgcgcaaa ggaacgaaaa agagattgcc 1140
107 gacattttcc gccaggcatt ctga
109 <210> SEQ ID NO: 4

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/037,677A

DATE: 04/22/2003

TIME: 14:28:34

Input Set : A:\GC560-D1-SEQLIST.txt

Output Set: N:\CRF4\04222003\J037677A.raw

```

110 <211> LENGTH: 387
111 <212> TYPE: PRT
112 <213> ORGANISM: Escherichia blattae
114 <400> SEQUENCE: 4
115 Met Ser Tyr Arg Met Phe Asp Tyr Leu Val Pro Asn Val Asn Phe Phe
116 1 5 10 15
117 Gly Pro Gly Ala Val Ser Val Val Gly Gln Arg Cys Gln Leu Leu Gly
118 20 25 30
119 Gly Lys Lys Ala Leu Leu Val Thr Asp Lys Gly Leu Arg Ala Ile Lys
120 35 40 45
121 Asp Gly Ala Val Asp Gln Thr Val Lys His Leu Lys Ala Ala Gly Ile
122 50 55 60
123 Glu Val Val Ile Phe Asp Gly Val Glu Pro Asn Pro Lys Asp Thr Asn
124 65 70 75 80
125 Val Leu Asp Gly Leu Ala Met Phe Arg Lys Glu Gln Cys Asp Met Ile
126 85 90 95
127 Ile Thr Val Gly Gly Gly Ser Pro His Asp Cys Gly Lys Gly Ile Gly
128 100 105 110
129 Ile Ala Ala Thr His Pro Gly Asp Leu Tyr Ser Tyr Ala Gly Ile Glu
130 115 120 125
131 Thr Leu Thr Asn Pro Leu Pro Pro Ile Ile Ala Val Asn Thr Thr Ala
132 130 135 140
133 Gly Thr Ala Ser Glu Val Thr Arg His Cys Val Leu Thr Asn Thr Lys
134 145 150 155 160
135 Thr Lys Val Lys Phe Val Ile Val Ser Trp Arg Asn Leu Pro Ser Val
136 165 170 175
137 Ser Ile Asn Asp Pro Leu Leu Met Ile Gly Lys Pro Ala Gly Leu Thr
138 180 185 190
139 Ala Ala Thr Gly Met Asp Ala Leu Thr His Ala Val Glu Ala Tyr Ile
140 195 200 205
141 Ser Lys Asp Ala Asn Pro Val Thr Asp Ala Ser Ala Ile Gln Ala Ile
142 210 215 220
143 Lys Leu Ile Ala Thr Asn Leu Arg Gln Ala Val Ala Leu Gly Thr Asn
144 225 230 235 240
145 Leu Lys Ala Arg Glu Asn Met Ala Cys Ala Ser Leu Leu Ala Gly Met
146 245 250 255
147 Ala Phe Asn Asn Ala Asn Leu Gly Tyr Val His Ala Met Ala His Gln
148 260 265 270
149 Leu Gly Gly Leu Tyr Asp Met Ala His Gly Val Ala Asn Ala Val Leu
150 275 280 285
151 Leu Pro His Val Cys Arg Tyr Asn Leu Ile Ala Asn Pro Glu Lys Phe
152 290 295 300
153 Ala Asp Ile Ala Thr Phe Met Gly Glu Asn Thr Thr Gly Leu Ser Thr
154 305 310 315 320
155 Met Asp Ala Ala Glu Leu Ala Ile Ser Ala Ile Ala Arg Leu Ser Lys
156 325 330 335
157 Asp Val Gly Ile Pro Gln His Leu Arg Glu Leu Gly Val Lys Glu Ala
158 340 345 350
159 Asp Phe Pro Tyr Met Ala Glu Met Ala Leu Lys Asp Gly Asn Ala Phe

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/037,677A

DATE: 04/22/2003

TIME: 14:28:34

Input Set : A:\GC560-D1-SEQLIST.txt

Output Set: N:\CRF4\04222003\J037677A.raw

```

160          355          360          365
161 Ser Asn Pro Arg Lys Gly Asn Glu Lys Glu Ile Ala Asp Ile Phe Arg
162          370          375          380
163 Gln Ala Phe
164 385
166 <210> SEQ ID NO: 5
167 <211> LENGTH: 1164
168 <212> TYPE: DNA
169 <213> ORGANISM: Escherichia blattae
171 <400> SEQUENCE: 5
172 atgagctatc gtatgtttga ttatctggtt ccaaattgtra acttcttttg cccgggcgcc      60
173 gtttctgttg ttggccagcg ctgccagctg ctggggggta aaaaagccct gctggtgacc      120
174 gataagggcc tgcgcgccat taaagacggt gctgtcgatc agaccgtgaa gcacctgaaa      180
175 gccgcccgtg ttgaggtggt cattttgcgac ggggtcgagc cgaaccgaa agacaccaac      240
176 gtgctcgacg gcctggccat gttccgtaaa gagcagtgcg acatgataat caccgtcggc      300
177 ggcggcagcc cgctcgactg cggtaaaggc attggtattg cggccacca cccgggtgat      360
178 ctgtacagct atgccggtat cgaaacactc accaaccgcg tgcgcgccat tattgcggtc      420
179 aacaccaccg ccgggaccgc cagcgaagtc acccgccact gcgtgctgac taacacaaaa      480
180 accaaagtaa aatttgtgat tgtcagctgg cgcaacctgc cttccgtctc cattaacgat      540
181 ccgctgctga tgatcggcaa gccgcgggg ctgaccgcg ccaccggtat ggatgccctg      600
182 acccacgcgg tagaggccta tatctccaaa gacgccaacc cggttaccga tgcctctgct      660
183 attcaggcca tcaaactgat tgccaccaac ttgcgccagg ccgtcgccct ggggaccaac      720
184 ctcaaagccc gtgaaaacat ggccctgcgc tctctgctgg ccgggatggc ctttaacaac      780
185 gccaacctgg gctatgttca cgccatggct caccagctgg gcggcctgta cgacatggcc      840
186 cacggggtgg cgaacgcggt cctgctgccc catgtctgcc gctataacct gattgccaac      900
187 ccggaataat ttgccgatat cgccaccttt atgggggaaa acaccaccg tctttccacc      960
188 atggacgcag cggagctggc catcagcgcc attgccgctc tgtctaaaga tgcgggatc      1020
189 ccgcagcacc tgcgtgaact gggggtaaaa gaggcgact tcccgatcat ggcagaaatg      1080
190 gccctgaaag acggcaacgc cttctctaac ccgcgcaaa ggaacgaaaa agagattgcc      1140
191 gacattttcc gccaggcatt ctga
193 <210> SEQ ID NO: 6
194 <211> LENGTH: 387
195 <212> TYPE: PRT
196 <213> ORGANISM: Escherichia blattae
198 <400> SEQUENCE: 6
199 Met Ser Tyr Arg Met Phe Asp Tyr Leu Val Pro Asn Val Asn Phe Phe
200 1          5          10          15
201 Gly Pro Gly Ala Val Ser Val Val Gly Gln Arg Cys Gln Leu Leu Gly
202          20          25          30
203 Gly Lys Lys Ala Leu Leu Val Thr Asp Lys Gly Leu Arg Ala Ile Lys
204          35          40          45
205 Asp Gly Ala Val Asp Gln Thr Val Lys His Leu Lys Ala Ala Gly Ile
206          50          55          60
207 Glu Val Val Ile Phe Asp Gly Val Glu Pro Asn Pro Lys Asp Thr Asn
208 65          70          75          80
209 Val Leu Asp Gly Leu Ala Met Phe Arg Lys Glu Gln Cys Asp Met Ile
210          85          90          95
211 Ile Thr Val Gly Gly Gly Ser Pro Leu Asp Cys Gly Lys Gly Ile Gly
212          100          105          110

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/037,677A

DATE: 04/22/2003

TIME: 14:28:34

Input Set : A:\GC560-D1-SEQLIST.txt

Output Set: N:\CRF4\04222003\J037677A.raw

```

213 Ile Ala Ala Thr His Pro Gly Asp Leu Tyr Ser Tyr Ala Gly Ile Glu
214      115      120      125
215 Thr Leu Thr Asn Pro Leu Pro Pro Ile Ile Ala Val Asn Thr Thr Ala
216      130      135      140
217 Gly Thr Ala Ser Glu Val Thr Arg His Cys Val Leu Thr Asn Thr Lys
218 145      150      155      160
219 Thr Lys Val Lys Phe Val Ile Val Ser Trp Arg Asn Leu Pro Ser Val
220      165      170      175
221 Ser Ile Asn Asp Pro Leu Leu Met Ile Gly Lys Pro Ala Gly Leu Thr
222      180      185      190
223 Ala Ala Thr Gly Met Asp Ala Leu Thr His Ala Val Glu Ala Tyr Ile
224      195      200      205
225 Ser Lys Asp Ala Asn Pro Val Thr Asp Ala Ser Ala Ile Gln Ala Ile
226      210      215      220
227 Lys Leu Ile Ala Thr Asn Leu Arg Gln Ala Val Ala Leu Gly Thr Asn
228 225      230      235      240
229 Leu Lys Ala Arg Glu Asn Met Ala Cys Ala Ser Leu Leu Ala Gly Met
230      245      250      255
231 Ala Phe Asn Asn Ala Asn Leu Gly Tyr Val His Ala Met Ala His Gln
232      260      265      270
233 Leu Gly Gly Leu Tyr Asp Met Ala His Gly Val Ala Asn Ala Val Leu
234      275      280      285
235 Leu Pro His Val Cys Arg Tyr Asn Leu Ile Ala Asn Pro Glu Lys Phe
236      290      295      300
237 Ala Asp Ile Ala Thr Phe Met Gly Glu Asn Thr Thr Gly Leu Ser Thr
238 305      310      315      320
239 Met Asp Ala Ala Glu Leu Ala Ile Ser Ala Ile Ala Arg Leu Ser Lys
240      325      330      335
241 Asp Val Gly Ile Pro Gln His Leu Arg Glu Leu Gly Val Lys Glu Ala
242      340      345      350
243 Asp Phe Pro Tyr Met Ala Glu Met Ala Leu Lys Asp Gly Asn Ala Phe
244      355      360      365
245 Ser Asn Pro Arg Lys Gly Asn Glu Lys Glu Ile Ala Asp Ile Phe Arg
246      370      375      380
247 Gln Ala Phe
248 385
250 <210> SEQ ID NO: 7
251 <211> LENGTH: 12
252 <212> TYPE: DNA
253 <213> ORGANISM: Artificial Sequence
255 <220> FEATURE:
256 <223> OTHER INFORMATION: wild type mutD gene
258 <400> SEQUENCE: 7
259 atgaccgcta tg
261 <210> SEQ ID NO: 8
262 <211> LENGTH: 11
263 <212> TYPE: DNA
264 <213> ORGANISM: Artificial Sequence
266 <220> FEATURE:

```

12

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/037,677A

DATE: 04/22/2003

TIME: 14:28:35

Input Set : A:\GC560-D1-SEQLIST.txt

Output Set: N:\CRF4\04222003\J037677A.raw